Nucleic Acids Research, 1996, Vol. 24, No. 2 303–310

Structure, function and expression of a murine homeobox protein AREC3, a homologue of *Drosophila sine oculis* gene product, and implication in development

Kiyoshi Kawakami*, Hiromi Ohto, Keiko Ikeda and Robert G. Roeder[1]

Department of Biology, Jichi Medical School, Minamikawachi, Kawachi, Tochigi, 329-04 Japan and [1]Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, New York, NY 10021, USA

Received September 29, 1995; Revised and Accepted November 28, 1995

GenBank accession nos D50416–D50418 (incl.)

ABSTRACT

The cDNA clones encoding ARE (Na,K-ATPase α1 subunit gene regulatory element) binding protein AREC3 were isolated from myoblast C2C12 cells and mouse skeletal muscle cDNA library. At least four alternatively spliced forms of AREC3 cDNA were identified. Sequence analysis indicates that AREC3 has an extensive homology with the *Drosophila sine oculis* gene product required for development of the entire visual system [Cheyette *et al.* (1994) *Neuron* 12, 977–996]. The homologous region including a homeodomain is required for specific DNA binding to ARE. A transactivation domain was identified in the C-terminal part of the AREC3 by reporter gene assays using GAL4–AREC3 fusion protein constructs. Immunohistochemistry revealed that AREC3 localized to the nucleus and cytoplasm of myoblast C2C12 cells, and the production of AREC3 is augmented during muscle differentiation. Western blot analysis indicated that the 115 kDa form of AREC3 protein is increased in the cytoplasmic extract, and the 67 kDa form is increased both in nuclear and cytoplasmic extracts of C2C12 cells during muscle differentiation.

INTRODUCTION

Na,K-ATPase is the enzyme responsible for maintaining the Na^+ and K^+ gradients across the cell membrane. The enzyme is composed of two subunits named α and β (1,2). The α subunit is the catalytic subunit in which ATP, cation and ouabain binding sites reside. At least three isoform genes α1, α2 and α3 have been identified (3). They are expressed in different tissues and are regulated under different developmental conditions (4). Each gene product exhibits different Na^+, K^+ and ouabain affinity (5,6). The Na,K-ATPase α1 subunit gene is expressed in all tissues and is most important for cellular homeostasis. We have analyzed the regulatory region of the gene in various cell lines and tissues using transient transfection or cell-free transcription systems (7,8). ARE was identified as a positive regulatory region common to various cell lines and tissues. More than seven ARE binding proteins are known, some of which are ubiquitous, and others tissue-specific (7,9). AREC3 protein was first identified as a cell-type specific ARE binding factor (7). In this study, we have isolated alternatively spliced forms of AREC3 cDNA from myoblast C2C12 cells and mouse skeletal muscle. We also identified the specific DNA-binding and transactivation domains, and analyzed the expression of the gene product in cultured cells and murine tissues. The implications for a developmental role of AREC3 are discussed.

MATERIALS AND METHODS

Screening and sequencing of cDNA

A partial cDNA clone encoding AREC3 was obtained from partial amino acid sequences of the purified AREC3 protein (Suzuki *et al.*, unpublished). The *Hin*dIII(547)–*Bst*EII (1580) fragment of human AREC3 cDNA (Suzuki *et al.*, unpublished) was used as a probe for screening the C2C12 cDNA library. About 1×10^6 plaques were screened, and 13 positive clones obtained. One of the longest clones (M18) and one of the shorter clones (M8), which has a different restriction map from that of M18, were subcloned into pSVSPORT (named pSVSPORTM18 and pSVSPORTM8, respectively) and sequenced. To obtain the full-length clone, we screened another library from mouse skeletal muscle (obtained from Clontech) using probes containing nt 1 to 869 of M18 and 154 to 322 of M8. We obtained five clones (f1, f2, f3, f4 and f14) that harbor the skeletal muscle-specific sequence fused to position 54 of M18 sequence from 2×10^6 plaques. The two overlapping clones of the most 5′-extended clone f2 and the most 3′-extended clone f39 were sequenced. Sequence homology was searched by FASTA (DDBJ). The accession numbers for cDNA sequences are D50416 (M18), D50417 (M8) and D50418 (SM).

* To whom correspondence should be addressed

RT–PCR

Skeletal muscle RNA was isolated from BALB/c mice (7 weeks old) using ISOGEN (Wako). Purification of polyA$^+$ RNA was by oligo dT-cellulose column. Transcription of polyA$^+$ RNA (2 μg) was carried out using 12.5 U AMV reverse transcriptase (Boehringer) with a primer of ATGCGCCTGAAAGTGGATGAGGAGACA (KK9411) in the presence of 26 U RNase inhibitor and 1 mM of each dNTP. PCR was performed in the presence of 0.2 mM dNTP and 0.2 μM of each primer. Thirty cycles of 94°C 45 s, 60°C 45 s and 72°C 2 min 30 s were performed using a Perkin Elmer thermal cycler. PCR primers used were 5′-AGTGGAGTTGTACCTGATCTGCCGCTC (KK9407), 5′-GCAGCTATCCGTTTACACTACTATTTGCTC (KK9410), 5′-CAATACACCGTCTCCTCGCCGTCCC (KK9412) and 5′-TGCGAGGGAGTGAGGGAGGGAGGAA (KK9415).

Northern hybridization

Mouse MTN blot was obtained from Clontech. *Xho*I–*Hin*dIII (318–869) fragment of pSVSPORTM18 was labeled by Megaprime (Amersham) and used as a probe. Hybridization was done using QuikHyb™ (Stratagene) according to the manufacturer's protocol.

GST fusion proteins

Various deletion proteins of the GST–AREC3 fusion were constructed as follows. For MB the *Sal*I–*Dra*I (–17–2486) fragment of pSVSPORTM18, containing the whole coding region, was blunt-ended with Klenow and ligated into the *Sma*I site of pGEX-3X. For MBNT, the MB construct was digested with *Hin*dIII (869) and *Eco*RI, blunt-ended and self-ligated. For MBCT, the *Hin*dIII–*Dra*I (869–2486) fragment of pSVSPORTM18 was blunt-ended and then ligated into the *Sma*I site of pGEX-3X. N-terminal deletions derived from MBNT were constructed as follows: for MBNTNΔ1, pSVSPORTM18 was cut with *Eco*47I (214), coupled with a *Bam*HI linker then digested with *Sac*I (531) and subcloned into *Bam*HI/*Sac*I-cut MBNT. For MBNTNΔ2, pSVSPORTM18 was cut with *Eco*47III (269–1948), coupled with a *Bam*HI linker, then cut with *Sac*I (531) and subcloned into *Bam*HI/*Sac*I-cut MBNT. For MBNTNΔ3, the *Xho*I–*Hin*dIII (318–869) fragment was blunt-ended and subcloned into the blunt-ended *Eco*RI site of pGEX-3X. For MBNTNΔ4, the *Bss*HII–*Hin*dIII (402–869) fragment of pSVSPORTM18 was blunt-ended, then subcloned into the blunt-ended *Eco*RI site of pGEX-3X. C-terminal deletions derived from MBNT were constructed as follows: for MBNTCΔ1 and MBNTCΔ2, the *Sac*I–*Pma*CI (531–792) fragment and the *Sac*I–*Mae*III (531–643) (blunt-ended) fragment of pSVSPORTM18, respectively, were subcloned into cut, blunt-ended *Eco*RI- and *Sac*I-digested MB. For MBNTCΔ3, MBNT was cut with *Msc*I (580) and blunt-ended, then digested with *Bam*HI and subcloned into *Bam*HI–*Eco*RI, blunt-ended digested pGEX-3X. For MBNTCΔ4, the *Sal*I–*Hin*II (–17–420) fragment of pSVSPORTM18 was blunt-ended and subcloned into *Sma*I digested pGEX-3X. For MBHD1, the *Rsa*I (434–697) fragment of pSVSPORTM18 was ligated into the blunt-ended *Eco*RI site of pGEX-3X. For MBHD2, the *Hin*1I (420–765) fragment of pSVSPORTM18 was blunt-ended, coupled with a *Bam*HI linker and subcloned into the *Bam*HI site of pGEX-3X. GST-fusion constructs of skeletal muscle type cDNA were constructed as follows: for SMNT, SMNTNΔ1, SMNTNΔ2 and SMNTNΔ3, cDNA clone f14 was digested with *Bal*31, blunt-ended, coupled with a *Bam*HI linker digested with *Hin*dIII (1096) and subcloned into *Bam*HI/*Hin*dIII-cut pKS. cDNAs harboring positions 64, 124, 172 and 310 to 1096 were excised with *Bam*HI and *Hin*cII, then subcloned into blunt-ended *Bam*HI/*Eco*RI digested pGEX-3X. GST fusion proteins were induced by adding IPTG, then purified on a gluthathione-Sepharose column (Pharmacia) following the manufacturer's protocol.

Gel retardation assays

Gel retardation assays were performed as described (10). ARE fragment (*Pvu*II–*Mlu*I fragment of *Atp1a1*) (7) was labeled with ^{32}P-dCTP and used as a probe. For competitors, C3WT: 5′-TCGAGCCGGTGTCAGGTTGCTCC and 5′-TCGAGGAGCAACCTGACACCGGC were annealed and C3MUT: 5′-TCGAGCCGGTGTGAGGTTGCTCC and 5′-TCGAGGAGCAACCTCACACCGGC were annealed.

Construction of GAL4 fusion protein and reporter gene assays

For GAL4F, GAL4NTCΔ4 and GAL4CT, the *Sal*I–*Dra*I (–17–2486), *Sal*I–*Bss*HII (–17–402) and *Hin*dIII–*Dra*I (869–2486) fragments from pSVSPORTM18 were blunt-ended and ligated into the *Sma*I site of pCMVΔGal4, which contains the GAL4 DNA-binding domain (1–147). For GAL4NTNΔ4, the *Bss*HII–*Hin*dIII (402–869) fragment from pSVSPORTM18 was blunt-ended, coupled with *Xba*I linker and ligated into *Xba*I site of pCMVΔGal4. For further dissection of the C terminal portion of AREC3, GAL4CT1 and GAL4CT2, the *Hin*dIII–*Pvu*II (869–1332) and *Pvu*II–*Apa*L1 (1332–1576) fragments from pSVSPORTM18 were blunt-ended and ligated into the *Sma*I site of pCMVΔGal4, while for GAL4CT3, the *Apa*L1–*Dra*I (1576–2486) fragment from pSVSPORTM18 was blunt-ended and ligated into the blunt-ended *Xba*I site of pCMVΔGal4. Reporter plasmid 1×UAS/CAT, which contains a GAL4 binding site in front of the HTLV-1 LTR promoter, was supplied by Dr Okuda. Transient transfection into C2C12 cells and CAT assays were performed as described (11).

Antibody preparation

The *Hpa*II–*Dra*I (1427–1918) and *Pvu*II (1007–1563) fragments of human AREC3 cDNA, for anti-AREC3 sera, nos 5 and 6 respectively, were fused to the *Eco*RI (blunt-ended) and *Sma*I sites of pGEX-3X, respectively. Fusion proteins were purified using a glutathione-Sepharose column (Pharmacia). The purified protein was mixed with Freund's complete adjuvant and injected into male rabbits three times in two weeks intervals. The antisera were absorbed with glutathione S-transferase.

Immunohistochemistry

Cells of C2C12 were grown in a 24-well culture dish in DMEM with 10% fetal bovine serum (FBS) or in DMEM with 2% horse serum as a differentiation medium. For immunostaining with the anti AREC3 antibody, cultures were rinsed with 1% FBS, washed with PBS, fixed with 2% paraformaldehyde in PBS for 60 min at room temperature and then permeabilized with PBS containing 0.2% Triton X-100 for 15 min at room temperature. Blocking was done using a labeled streptavidin biotin (LSAB) kit from DAKO

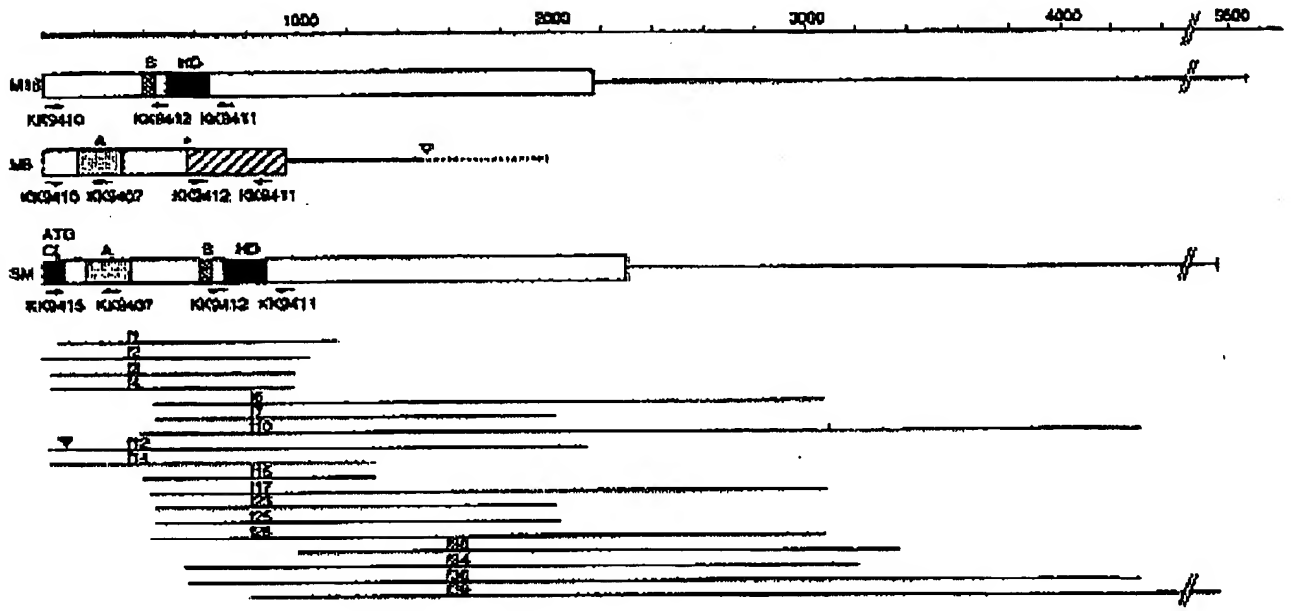

Figure 1. Structural feature of cDNA clones of AREC3. Homeodomain (HD) and specific exons A, B and C are indicated. Asterisks indicate the position of frame shift in M8 sequence. Open arrowhead shows the position of diverged sequence in M8. Closed arrowhead indicates the position of f12 to which the same sequence as for M18 is fused. The positions of oligonucleotides used for RT–PCR are indicated by arrows. The region of cDNA clones from skeletal muscle library are indicated.

and incubating with the absorbed rabbit anti-AREC3 serum 6 at 1:500 dilution and with anti-Sp1 antibody (Santa Cruz Biotech.) at 1:50 dilution for 30 min. Visualization of AREC3 was accomplished with the LSAB kit.

Western blot analysis

Nuclear extracts from C2C12 cells in growing medium and in differentiation medium were prepared according to the method described in Kawakami *et al.* (12), modified by adding aprotinin (14 μg/ml), pepstatin A (0.1 μg/ml), leupeptin (0.1 μg/ml), antipain (0.1 μg/ml) and soybean trypsin inhibitor (2 μg/ml) in buffer B. The cytoplasmic supernatants of the nuclear pellet were pooled as cytoplasmic extracts. Protein (20 μg) from the nuclear and cytoplasmic extracts was resolved by 9% polyacrylamide–SDS gel electrophoresis. Proteins were transferred to Hybond-ECL membrane (Amersham) and analyzed with anti-AREC3 serum 5 at 1:3000 dilution using the ECL Western blotting analysis system (Amersham).

RESULTS

Cloning of mouse AREC3 cDNA

To understand the structure and function of the ARE binding factor AREC3, we obtained three alternatively spliced cDNA clones: M8 and M18 from the cDNA library of C2C12 cells, in which AREC3 was known to be produced by gel retardation assay (unpublished result), and SM from mouse skeletal muscle. As shown in Figure 1, M8 contains the specific exon A which does not appear in M18, while M18 contains the specific exon B. M8 has no exon B (resulting in the frame shift at position 588) and codes a totally different amino acid sequence from amino acid position 196 alanine. Among 18 clones from plaques of mouse skeletal muscle cDNA library that hybridize with the 5′ end portion of M18, five (f1, f2, f3, f4 and f14) extended beyond position 54 of M18, nested and contained the muscle-specific sequence (indicated as C in Fig. 1). The ATG codon is found at position 62, after the termination codon at the position 32 in frame. The other twelve clones overlap with M18. The structure of the combined nucleotide sequence of two overlapping skeletal muscle clones (f2 and f39, which covers from 1 to 1042 and from 819 to 5566, respectively) is shown as SM in Figure 1. We found no sequence difference from that of the M18 except two T residues missing at position 3267 in SM (corresponding to 3040 in M18) in the 3′ non-coding sequence. One of the 18 clones (f12, fourth alternatively spliced form) is different in that this type contains A and B exons and extends to nt 8 of M18. The homeodomain is found in the central portion of M18 and SM (Fig. 1).

To confirm that these alternatively spliced forms are relevant *in vivo*, we tried to identify these molecules by RT–PCR using mouse skeletal muscle poly A$^+$ RNA (Fig. 2). For this purpose, we made a set of PCR primers flanking the alternatively spliced exons A and B (KK9410 and KK9412). Control PCR using cDNA clones of pSVSPORTM18 (lane 1), pSVSPORTM8 (lane 2) and pSVSPORTf12 (lane 3) as templates indicates the expected position of PCR products. We detected PCR product corresponding to M18 by RT–PCR (lane 6). After we increased the PCR cycles, the products corresponding to M18 and f12 were observed (lane 7). To confirm the presence of skeletal muscle type mRNA, we tested another set of primers in skeletal muscle-specific sequence corresponding to nt 22 to 46 of SM (KK9415) and in the exon A (KK9407). PCR product corresponding to the

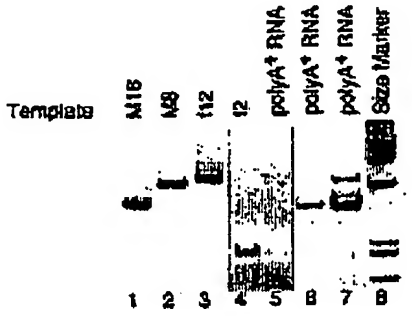

Figure 2. Reverse transcriptase PCR analysis of mouse skeletal muscle polyA+ RNA. Control PCR reactions using cDNA clones of pSVSPORTM18 (lane 1), pSVSPORTM8 (lane 2), pSVSPORTf12 (lane 3) and pSVSPORT f2 (lane 4) as templates. PolyA+ RNA was reverse transcribed with a primer of KK9411 and used for PCR templated reactions (lanes 5–7). PCR primers were KK9410 and KK9412 (lanes 1–3, 6 and 7), KK9407 and KK9415 (lane 4 and 5). Thirty thermo-cycles were performed except for lane 7 in which 40 cycles were done. DNA size markers (φX174 *Hae*III digest) are shown (lane 8).

muscle-specific type cDNA was observed (lane 5) at the same mobility as the control PCR using PSVSPORTf2 as template (lane 4). Omission of one of the primers or template cDNA gave no PCR product in any PCR reaction (data not shown). These results indicate that the M8 type mRNA does not exist in muscle, while muscle-specific type mRNA (containing A, B and C exons) and M18 type mRNA (containing B exon only) are the major species in muscle. However, the other type (f12), containing A and B but not C exons, also exists as a minor species.

Homology with *Drosophila sine oculis* protein

By searching the DDBJ data library, the nucleotide sequence of AREC3 cDNA was revealed to have extensive homology with the cDNA of *sine oculis* gene product from *Drosophila* required for the development of the entire visual system (13,14). Figure 3 shows the alignment of the amino acid sequence of skeletal muscle type AREC3 and that of *Drosophila sine oculis* gene product. The extensively homologous region spans from valine 106 to arginine 273 of the skeletal muscle type AREC3. The homeodomain (aspartic acid 216 to alanine 275) is almost included in the C-terminal portion of this region. The overall homology of the region is 68%. The minimal essential region required for specific binding is included in this extensively homologous region (see below). This finding raised the possibility that AREC3 is involved not only in muscle differentiation but also in visual system development.

Tissue distribution of AREC3 mRNA

The tissue distribution of AREC3 mRNA was analyzed by Northern blot analysis using a probe of *Xho*I–*Hin*dIII (318–869) fragment of pSVSPORTM18. Figure 4 shows that AREC3 is expressed mainly in skeletal muscle (lane 6) and weakly in heart (lane 1). The size of the mRNA is ~7.0 kb. Faint signals of the same size were observed in lung and kidney with longer exposure (data not shown). In brain, mRNA of ~8.0 kb in length was observed (lane 2).

Identification of specific DNA-binding domain

To identify the DNA-binding domain of the AREC3 molecule, we made GST-fusion proteins harboring various portions of the AREC3 (Fig. 5A). Proteins were purified by glutathione-Sepharose column chromatography and analyzed by gel retardation assays using the ARE fragment as probe.

First, MBNT and MBCT, which contain the N-terminal half of MB protein (encoded by M18 cDNA) and the C-terminal half of MB protein, respectively, were tested for DNA-binding activity in the gel retardation assay. As shown in Figure 5B, a retarded complex was observed for MBNT (lane 2) but not for MBCT (lane 3). This indicates that the DNA-binding activity of AREC3 resides in the N-terminal portion of AREC3. Skeletal muscle type SMNT also retains the DNA-binding activity (Fig. 5D, lane 1). To map precisely the DNA-binding domain in the MBNT, we made a series of N- and C-terminal deletion mutations and tested their binding activity. Retarded complexes were observed for MBNTNΔ1 to 4 (Fig. 5B, lanes 4–7) and MBNTCΔ1 to 3 (lanes 8–10), but not in MBNTCΔ4 (lane 11). These results indicate that the DNA-binding domain of AREC3 resides in the region from arginine 135 to leucine 193 (position for M18), which contains the homeodomain. This was confirmed by demonstrating that the other two constructs of MBHD1 and MBHD2 showed DNA binding activity (Fig. 5B, lanes 12,13). To verify that the observed binding activity is specific to the AREC3 binding sequence, we

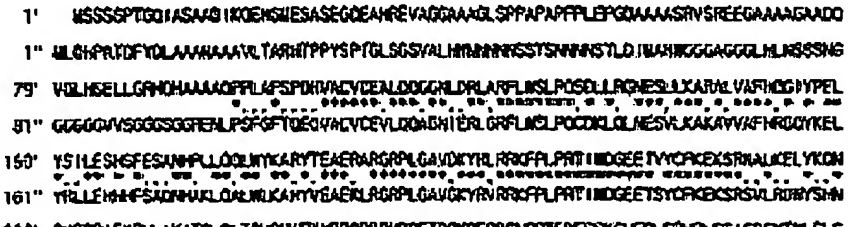

Figure 3. Sequence homology and alignment of AREC3 and *Drosophila sine oculis* protein. Skeletal muscle-type AREC3 amino acid sequence (upper row) and that of *Drosophila sine oculis* protein (lower row) are aligned. Conserved amino acids are shown by asterisks, while conservative changes are indicated by dots. Sequence from 319 of AREC3 and that from 321 of *Drosophila sine oculis* protein to C-terminal are omitted.

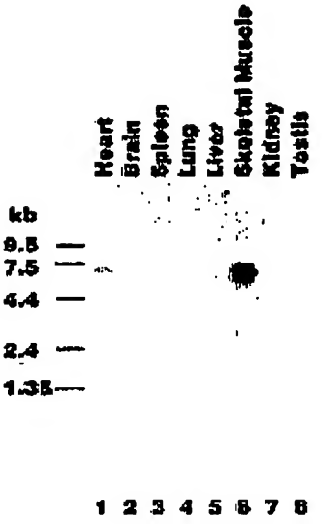

Figure 4. Northern hybridization of RNA from mouse tissues. PolyA$^+$ RNAs from mouse heart (lane 1), brain (lane 2), spleen (lane 3), lung (lane 4), liver (lane 5), skeletal muscle (lane 6), kidney (lane 7) and testis (lane 8) were applied. Positions of RNA size markers are shown.

examined the effect of the competitor C3WT, which covers the C3 binding sequence. As a control, we also tested C3MUT which contains one point mutation of the C3WT oligonucleotide. The competition experiment indicated that C3WT competed with MBNT complex formation (Fig. 5C, lanes 4–6), MBNTNΔ1 (lanes 7–9), MBNTNΔ2 (lanes 10–12), MBNTCΔ1 (lanes 19–21) and MBNTCΔ2 (lanes 22–24), using ARE probe as C3 formation in HeLa nuclear extract (lanes 1–3), but did not compete with MBNTNΔ3 (lanes 13–15), MBNTNΔ4 (lanes 16–18), MBNTCΔ3 (lanes 25–27), MBHD1 (lanes 28–30) and MBHD2 (lanes 31–33). These results suggest that the minimal essential region for the sequence-specific binding of AREC3 is from leucine 91 to aspartic acid 215, and that the homeodomain has a non-specific binding activity or one of other sequence specificity.

To know whether the region encoded by the SM type cDNA (exons C and A) has any effect on the activity or specificity of the DNA binding, we made a series of N-terminal deletions SMNTNΔ1, SMNTNΔ2 and SMNTNΔ3. All the constructs showed complex formation and the formations were competed by the C3WT and not by the C3MUT (Fig. 5D). The results indicate that these exons have no effect on the activity or specificity of DNA binding.

Identification of transcriptional activation domain

To know whether AREC3 has a transactivation domain, we made GAL4 fusion proteins containing the whole coding region of M18 (GAL4F), N-terminal (GAL4NTCΔ4), central (GAL4NTNΔ4) and C-terminal (GAL4CT) portions of AREC3, and tested the effect on HTLV1 LTR promoter containing a GAL4 binding site in the upstream region. As shown in Figure 6, the fusion construct containing the whole coding region (GAL4F) activated the promoter 3.7-fold and the C-terminal portion (GAL4CT) activated the promoter 9.3- or 21.1- fold compared with the GAL4 DNA-binding domain. This indicates that AREC3 has a transactivation domain in the C-terminal region. For precise mapping of the domain, we further dissected the C-terminal portion into three parts as shown in Figure 6. GAL4CT1 and GAL4CT2 showed little activation (2.0- and 0.7-fold), while GAL4CT3 exhibited the activation of 16.6-fold. From these results, we concluded that AREC3 has a transactivation domain in its C-terminal portion (526 alanine to 719 leucine).

Antibody production and specificity

To analyze the AREC3 protein distribution in various cells and tissues, we prepared rabbit polyclonal antibodies to human AREC3. GST-fusion protein from valine 477 to leucine 611 (serum 5) and leucine 337 to threonine 521 (serum 6) was used as antigen, corresponding to valine 585 to leucine 719 and leucine 445 to threonine 629 in mouse SM type AREC3, respectively. The resulting serum was tested for specificity in gel retardation assays. Formation of the typical C3 complex in HeLa nuclear extract was abolished and a slower migrating complex appeared in the presence of anti AREC3 serum 5, while preimmune serum gave no effect. The anti AREC3 serum 6 showed essentially the same results (data not shown). This confirms that our human cDNA clone corresponds to the HeLa C3 factor. The anti AREC3 sera also abolished the formation of C3 complex in nuclear extracts from mouse BALB/c-3T3 cells (data not shown), indicating the cross reactivity of this serum with the mouse AREC3.

Localization of AREC3 in cultured cells

To obtain insight into the involvement of AREC3 in muscle differentiation, we performed immunostaining of C2C12 myoblast cells with the anti AREC3 serum. As shown in Figure 7A, specific staining was observed both in the nucleus and the cytoplasm of growing C2C12 cells. The distribution of AREC3 in the cytoplasm looks particulate. The control experiment using anti-Sp1 antibody showed no particulate distribution in the cytoplasm (Fig. 7D). After 8 or 12 days in differentiation medium, muscle differentiated multi-nucleated cells were strongly stained with the serum, indicating that AREC3 is induced during muscle differentiation (Fig. 7B and C).

The cytoplasmic distribution and the induction of AREC3 during muscle differentiation were confirmed by Western blot analysis (Fig. 7E). Bands (4–5) were identified both in the nuclear (lanes 1,2) and cytoplasmic extracts (lanes 3,4). The 67 kDa form of the protein was increased in both nuclear and cytoplasmic extracts from cells in the differentiation medium (lanes 2,4). In contrast, the 115 kDa form of the protein was increased only in the cytoplasmic extract from cells in the differentiation medium (lane 4).

DISCUSSION

Identification of AREC3 cDNA clones

We identified one of the tissue-specific ARE binding factor AREC3. The specific antiserum against bacterially expressed human AREC3 recognized the HeLa cell C3 factor in gel retardation assays. The *bona fide* AREC3 cDNA (Suzuki *et al.* unpublished) was used for obtaining the murine homologue of AREC3.

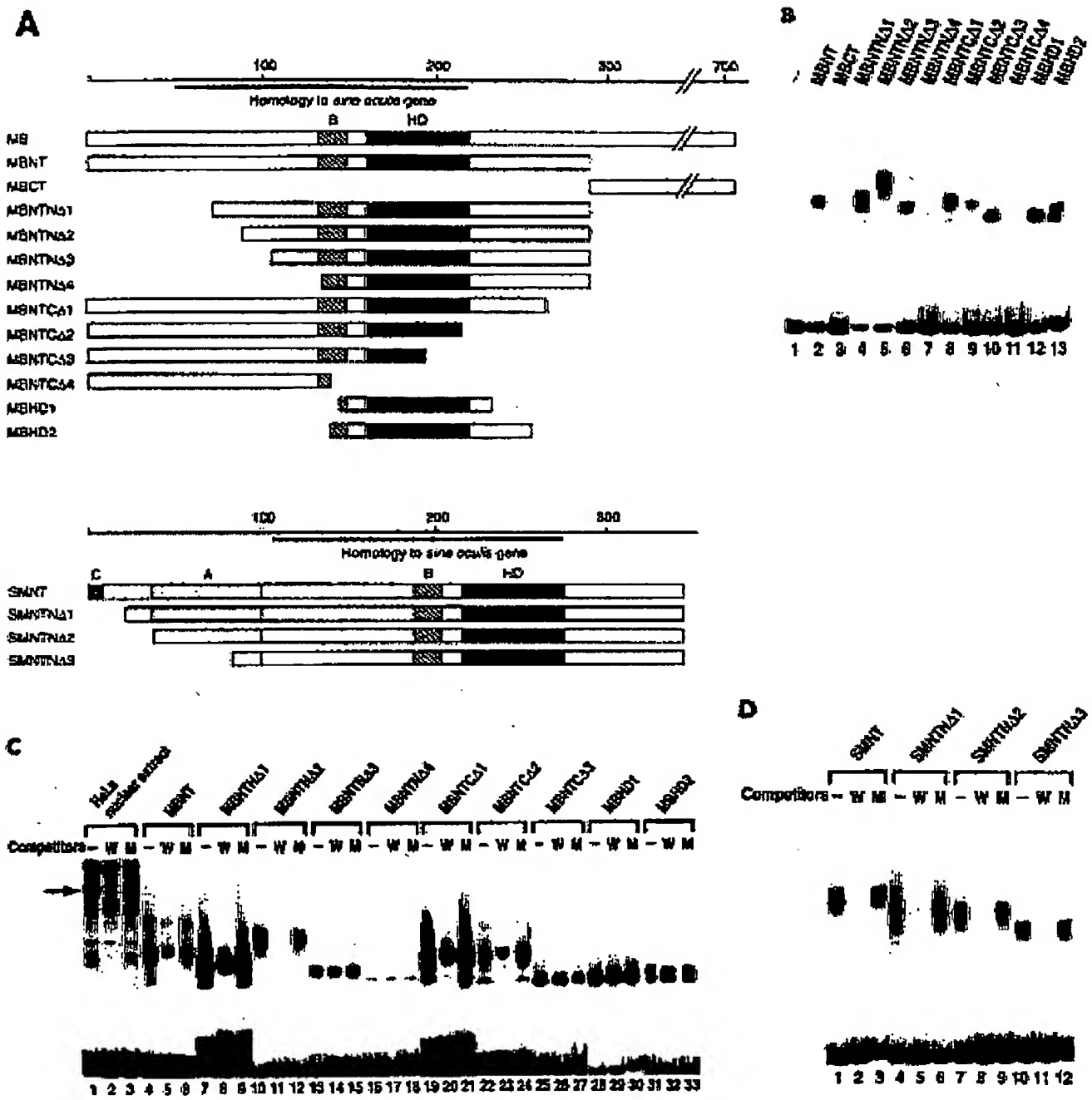

Figure 5. DNA-binding domain of AREC3. (A) Deletion constructs of GST–AREC3 fusion proteins. The regions of each deletion mutations are indicated. Scales indicate the amino acid residues. The region homologous to *Drosophila sine oculis* is shown. (B) Gel retardation assay of GST–AREC3 fusion proteins. MBNT, 45 ng (lanes 2); 120 ng MBCT (lane 3); 1.4 ng MBNTNΔ1 (lane 4); 3.0 ng MBNTNΔ2 (lane 5); 1.1 ng MBNTNΔ3 (lane 6); 6.3 ng MBNTNΔ4 (lane 7); 8.0 ng MBNTCΔ1 (lane 8); 13 ng MBNTCΔ2 (lane 9); 9.2 ng MBNTCΔ3 (lane 10); 17.3 ng MBNTCΔ4 (lane 11); 75 ng MBHD1 (lane 12); 26 ng MBHD2 (lane 13). The arrow indicates the position of C3 in HeLa nuclear extract. (C) Competition assay of GST–AREC3 fusion proteins. HeLa protein nuclear extract, 5 µg (lanes 1–3); 45 ng MBNT (lanes 4–6); 2.3 ng MBNTNΔ1 (lanes 7–9); 6.0 ng MBNTNΔ2 (lanes 10–12); 1.1 ng MBNTNΔ3 (lanes 13–15); 6.3 ng MBNTNΔ4 (lanes 16–18); 10.0 ng MBNTCΔ1 (lanes 19–21); 13 ng MBNTCΔ2 (lanes 22–24); 9.2 ng MBNTCΔ3 (lanes 25–27); 150 ng MBHD1 (lanes 28–30); 26 ng MBHD2 (lanes 31–33). Competitors, 1 pmol, C3WT (lanes 2,5,8,11,14,17,20,23,26,29 and 32) and C3MUT (lanes 3,6,9,12,15,18,21,24,27,30 and 33). (D) Competition assays of skeletal muscle type AREC3 protein. SMNT,4.0 ng (lanes 1–3); 4.5 ng SMNTΔ1 (lanes 4–6); 6.9 ng SMNTΔ2 (lanes 7–9); 10.1 ng SMNTΔ3 (lanes 10–12) were added. Competitors, 1 pmol (~60-fold molar excess to the probe) of C3WT (lanes 2, 5, 8 and 11) and C3MUT (lanes 3, 6, 9 and 12) were added.

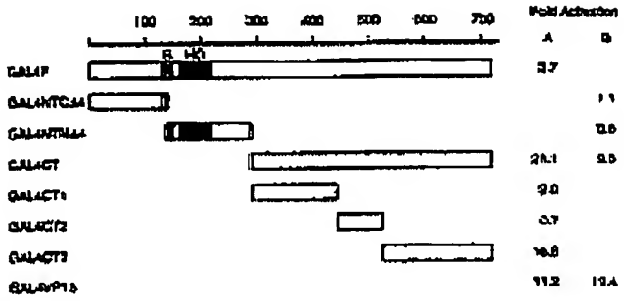

Figure 6. Transactivation domain of AREC3. The region in each construct of GAL4–AREC3 fusion proteins and the transactivation activities (compared with the CAT activity of GAL4 DNA binding domain) of each construct are shown. Two separate sets of experiments were performed.

Four types of alternatively spliced mRNA forms were obtained. At least three of them were confirmed to exist in adult skeletal muscle by RT–PCR. Furthermore, we also determined each of the alternatively spliced exon (A, B and C) in the mouse genomic sequences (Kawakami, unpublished observation), eliminating the possibility of cloning artifacts. Skeletal muscle-type mRNA encodes a characteristic protein in its N-terminal portion which contains an alanine tract. A similar structural feature is also observed in *Drosophila sine oculis* protein in that a homoglycine tract is observed in the N-terminal portion.

Homeodomain-containing protein is known to be involved in many developmental processes. *Atp1a1* is known to be regulated during development in various tissues (4). One of the factors, AREB6, which binds to the regulatory region ARE of Na,K-ATPase, also contains a homeodomain between the two zinc-finger clusters (9,15). The observations that the AREC3 is produced in restricted regions and at a specific time point during development (see below) suggest that AREC3 functions at a certain developmental stage, when *Atp1a1* is regulated. The repertoire of transcription factors regulating *Atp1a1* changes among cell types and tissues, and during the cell cycle (12). AREC3 might regulate the gene mainly in the early developmental stage *in vivo*.

Domain structure

AREC3 contains not only a homeodomain but also a newly identified motif of a region highly homologous to *Drosophila sine oculis* protein. Part of this region is involved in the specific binding activity of AREC3. Gel retardation analyses of various deletion proteins of AREC3 revealed that the minimal essential region for specific binding exists between leucine 91 and aspartic acid 215. The homeodomain itself shows DNA binding activity with different specificity. This situation is similar to the POU homeodomain protein in which both the POU-specific domain and the homeodomain are necessary for the specific binding (16,17). The AREC3 binding sequence GGNGNCNGGTTGC (7) is not homologous to other homeodomain proteins. The binding sequences of many homeodomain proteins include TAAT core motif (18). Other binding sequences such as GNNCACTCAAG of thyroid nuclear factor 1 (19) or TCACGCNTGA of *Pax* paired domain recognition sequence (20) do not fit the AREC3 binding sequence.

AREC3 also contains the transactivation domain in the region from alanine 526 to the C-terminus leucine 719 although we did

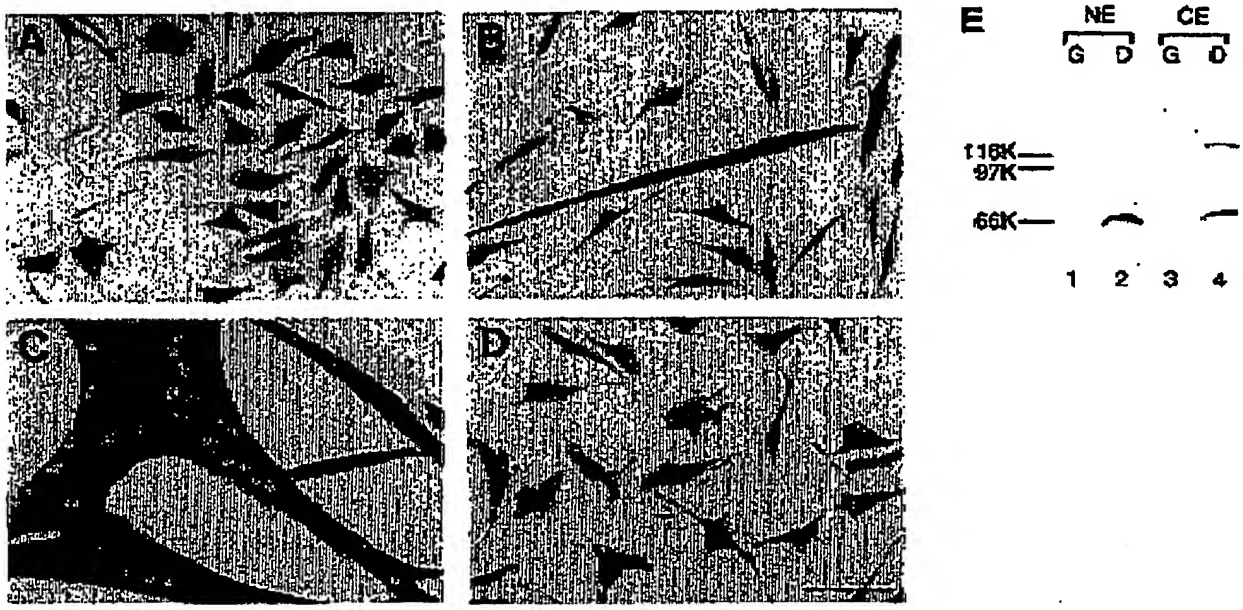

Figure 7. Immunohistochemistry of C2C12 cells and Western blot analysis of nuclear and cytoplasmic extracts with anti-AREC3 serum 6. (A) In the growing medium. (B) 8 days after in the differentiation medium. (C) 12 days after in the differentiation medium. (D) Stained with anti-Sp1 in the growing medium. Scale bar, 100 μm. (E) Western blot analysis with anti-AREC3 serum 5. Nuclear extracts from C2C12 cells in growth (lane 1) and in differentiation (lane 2) medium; cytoplasmic extracts from cells in growth (lane 3) and in differentiation (lane 4) medium were analyzed. Positions of size markers are shown.

not find any typical structure of the activation domain, such as Gln-rich, Pro-rich or acidic, in the region. These two features (specific DNA-binding and transactivation activity) strongly suggest that AREC3 functions as a transcription factor.

The role of AREC3 in development

The extensive homology of AREC3 with the *sine oculis* protein directly indicates the essential role for the protein during development of visual system. The other two *sine oculis*-related murine homeobox genes *Six1* and *Six2*, which are expressed during the development of limb tendons but not in retina, were recently identified (21). Although AREC3 has less similarity with *sine oculis* protein than *Six1* and *Six2* do, we observed the specific expression of AREC3 in retina by *in situ* hybridization and immunohistochemistry (Ohto *et al.*, unpublished observation). This suggests that the AREC3 is a functional homologue of *sine oculis*.

In contrast with retina, no production of AREC3 was observed in adult muscle, but it is produced transiently in embryonic or early stages after birth (Takizawa *et al.*, unpublished observation). It was also noted that the restricted distribution of AREC3 is observed in distal tubule of the kidney (Takizawa *et al.*, unpublished observation). These observations strongly suggest that AREC3 is important for the maintenance of retina, muscle and kidney during development.

ACKNOWLEDGEMENTS

This work was supported in part by grants from Ministry of Education, Science and Culture of Japan and from the Ciba Geigy Science Promotion Foundation and by the Science Research Promotion Fund from the Japan Private School Promotion Foundation. We thank Dr Takashi Obinata for providing a cDNA library from C2C12 cells, Drs Ken Yanagisawa and Shin-ichi Tominaga for preparation of BALB/c-3T3 cells, Drs Toshihiro Takizawa and Takuma Saito for advice in immunostaining, Dr Makoto Kobayashi for discussions and Mss Yuko Watanabe and Kuniko Takase for technical assistance.

REFERENCES

1 Sweadner,K.J. (1989) *Biochim. Biophys. Acta*, 988, 185–220.
2 Lingrel,J.B., Orlowski,J., Shull,M.M. and Price,E.M. (1990) Molecular genetics of Na,K-ATPase. In *Progress in Nucleic Acid Research and Molecular Biology*, Cohn,W.E. and Moldave,K. (eds), Academic, San Diego, Vol. 38, pp 37–89.
3 Shull,G.E., Greeb,J. and Lingrel,J.B. (1986) *Biochemistry*, 25, 8125–8132.
4 Orlowski,J. and Lingrel,J.B. (1988) *J. Biol. Chem.*, 263, 10436–10442.
5 Jewell,E.A. and Lingrel,J.B. (1991) *J. Biol. Chem.*, 266, 16925–16930.
6 Munzer,J.S., Daly,S.E., Jewell-Motz,E.A., Lingrel,J.B. and Bloszein,R. (1994) *J. Biol. Chem.*, 269, 16668–16676.
7 Suzuki-Yagawa,Y., Kawakami,K. and Nagano,K. (1992) *Mol. Cell. Biol.*, 12, 4046–4055.
8 Kobayashi,M. and Kawakami,K. (1995) *Nucleic Acids Res.*, 23, 2848–2855.
9 Watanabe, Y., Kawakami, K., Hirayama, Y. and Nagano, K. (1993) *J. Biochem. Tokyo*, 114, 849–855.
10 Kawakami,K., Scheidereit,C. and Roeder,R.G. (1988) *Proc. Natl Acad. Sci. USA*, 85, 4700–4704.
11 Ikeda,K., Nagano,K. and Kawakami,K. (1993) *Eur.J. Biochem.*, 218, 195–204.
12 Kawakami,K., Yanagisawa,K., Watanabe,Y., Tominaga,S. and Nagano,K. (1993) *FEBS Lett.*, 335, 251–254.
13 Cheyette,B.N.R., Green,P.J., Martin,K., Garren,H., Hartenstein,V. and Zipursky,S.L. (1994) *Neuron*, 12, 977–996.
14 Serikaku,M.A. and O'Tousa,J.E. (1994) *Genetics*, 138, 1137–1150.
15 Ikeda,K. and Kawakami,K. (1995) *Eur. J. Biochem.*, 233, 73–82.
16 Ingraham,H.A., Flynn,S.E., Voss,J.W., Albert,V.R., Kapiloff,M.S., Wilson,L. and Rosenfeld,M.G. (1990) *Cell*, 61, 1021–1033.
17 Sturm,R.A. and Herr,W. (1988) *Nature*, 336, 601–604.
18 Hayashi,S. and Scott,M.P. (1990) *Cell*, 63, 883–894.
19 Guazzi,S., Price,M., De Felice,M., Damante,G., Mattei,M.-G. and Di Lauro,R. (1990) *EMBO J.*, 9, 3631–3639.
20 Epstein,J., Cai,J., Glaser,T., Jepeal,L. and Maas,R. (1994) *J. Biol. Chem.*, 269, 8355–8361.
21 Oliver,G., Wehr,R., Jenkins,N.A., Copeland,N.G., Cheyette,B.N.R., Hartenstein,V., Zipursky,S.L. and Gruss,P. (1995) *Development*, 121, 693–705.